

Supporting Information

Structural distributions from single-molecule measurements as a tool for molecular mechanics

Jeffrey A. Hanson^{†‡}, Jason Brokaw[†], Carl C. Hayden[§], Jih-Wei Chu^{‡}, and Haw Yang^{†‡*}*

Department of [†]Chemistry and [‡]Chemical Engineering, the University of California at Berkeley,
Berkeley, CA 94720

[§]Combustion Research Facility, Sandia National Laboratories, P.O. Box 969, Livermore, CA 94551

[‡]Department of Chemistry, Princeton University, Princeton, NJ 08550

E-mail: hawyang@princeton.edu (HY), jwchu@berkeley.edu (JWC), cchayde@sandia.gov (CCH)

Contents

I. Single-molecule data selection criteria.....	3
II. Resolving power of the maximum-information method.....	5
III. Single-molecule spectra and lifetime distributions	7
IV. Single-molecule polarization measurement of dye randomization	9
V. Steady-State FRET in trifluoroethanol	11
VI. The freely jointed chain model.....	12
VII. The self-avoiding chain model.....	12
VIII. The worm-like chain model	13
IX. References	16

I. Single-molecule data selection criteria

Data selection is one of the most important steps in performing rigorous single-molecule experiments. The challenge for the experimentalist is objective elimination of those trajectories in which unwanted photo-physics (such as triplet-state blinking of the fluorophores) or multiple labeled molecules in the detection volume contribute to changes in emissive state that would otherwise be interpreted as conformational dynamics in single-molecule FRET experiments. This type of analysis is critical to single-molecule FRET experiments yet frequently performed manually, which can result in biased and inconsistent results. Furthermore, due to lack of rigorous criteria for well-behaved single-molecule trajectories, this essential step is almost never discussed in publications, which can result in wide variations in experimental results since this type of manual data selection is highly user dependent. In order to standardize data selection, a set of criteria has been implemented to reduce the variability arising from an entirely manual selection of FRET trajectories. Though it is impossible to completely automate trajectory selection, implementing this type of automated analysis as a precursor to manual selection will improve transparency and reproducibility in single-molecule data analysis. Criteria for semi-automated single-molecule data selection are detailed below.

The first criterion is that each trajectory has a signal-to-noise ratio in both donor and acceptor channels of greater than 5:1 (details discussed in the Supplementary Information in Ref. [1]). Moreover, single-molecule data with a poor signal-to-noise ratio will have a relatively slow time-resolution when analyzed according to the maximum information method (MIM) and will result in single-molecule probability density functions (PDFs) created from raw-data with a broad range of time-resolutions. Inclusion of this low quality data in subsequent analysis would complicate, if not completely invalidate, the interpretation of single-molecule PDFs as a function of mean time-resolution as presented previously [1].

Subsequent criteria used in automated trajectory selection rely on statistical tests derived for the detection auto- and cross-correlation in time-series [2, 3]. These tests are used to detect deviations from ideal behavior in single-molecule trajectories. Here, change-point analysis [4] is used to detect discreet changes in dye emission in order to break the single-molecule trajectory into three regions: in region 1 the dyes are undergoing energy transfer and both are emitting photons, in region 2 the acceptor dye has photo-bleached but the donor is still emitting and in region 3 both dyes have photo-bleached and only background counts are detected in each channel. Though only region 1 contains distance information

from FRET, regions 2 and 3 both contain important information which is used for calibrating the measured distances in the MIM method [5]. After the acceptor dye has bleached in single-molecule FRET experiments, the intensity in each channel is expected to be constant in both regions 2 and 3. Typically deviations from ideal behavior arise from triplet-state blinking in region 2 or additional bleaching events arising from multiple labeled molecules in the detection volume, both of which are undesirable since they could lead to miscalibration of the measured distances in region 1. In the ideal case, intensity auto-correlation functions for both donor and acceptor channels in regions 2 and 3 are expected to show no significant intensity auto-correlation. A statistical test for auto-correlation [2] is applied separately to each channel in regions 2 and 3 and those trajectories displaying significant auto-correlation at 95% confidence are removed from the data set and not included in subsequent data selection or analysis. In region 2, each channel is binned at 50 ms while in region 3 each channel is binned at 200 ms. These tests are intended to remove trajectories with large changes in emissive state such as blinking or bleaching of additional dyes in the detection volume, both of which would cause miscalibration in the data analysis parameters for MIM analysis.

An intensity cross-correlation analysis is used to assess the validity of the trajectory in region 1 [2]. Since the emission of the dyes is expected to be uncorrelated or negatively correlated, those trajectories with positive correlation to 95% confidence are removed from the data set. Here positive correlation arises from triplet-state blinking of the acceptor dye, which leads to erroneous distance calculations. Cross-correlation tests in region 1 are performed on data binned at 2 ms.

Another criterion used in data selection relies on a unique feature of the MIM analysis method. As mentioned above, regions 2 and 3 of each single-molecule trajectory are used to calibrate the distances measured in region 1 [5]. This is necessary since differences in experimental collection efficiencies from each channel as well as differences in intrinsic quantum yield between the fluorophores results in different intensities detected for each fluorophore. This correction is analogous to the gamma factor correction commonly applied to single molecule efficiency measurements.[6] In the MIM method, the calibration of donor and acceptor intensities is done on a molecule-by-molecule basis, which allows an equivalent to the gamma factor to be calculated for each molecule as I_A^0/I_D^0 , where I_D^0 is the average intensity of donor emission in the absence of acceptor and I_A^0 is the interpolated acceptor intensity at 100% energy transfer [5]. This molecule-by-molecule analysis reveals that the gamma factors are normally distributed rather than truly constant as is commonly assumed in conventional FRET analysis and is a necessary assumption in single-molecule diffusion type experiments. This is expected to arise from gradual drift in focus and instrumental alignment during data collection as well as error in the spot finding algorithm used to find the centroid of each molecule. Application of a constant gamma factor to

a single-molecule data set will cause calibration errors to manifest themselves as an artificially broadened distribution.

The I_A^0/I_D^0 ratio calculated in MIM analysis is used as a data selection since outliers are indicative of multiple chromophores in the detection volume or drastic focus and spot finding errors. In this case any molecule with an I_A^0/I_D^0 ratio lying more than three standard deviations from the mean is removed from subsequent data analysis.

II. Resolving power of the maximum-information method

It is important to characterize the resolving power of the data analysis approach, given that a potentially multimodal distribution has been observed. To investigate the intrinsic resolution of the single-molecule data analysis methods utilized in the present work, data sets of known composition were simulated by creating artificial, *in silico* mixtures of experimental poly-L-proline trajectories. Figure S1 shows the results of this exercise for both the maximum information method (MIM) and constant time binning (CTB) analysis as discussed in the main text. Binary mixtures, created by combining data from only two single-molecule trajectories, were constructed to investigate the performance of the maximum entropy deconvolution algorithm used to remove noise from probability density functions (PDFs) (Figure S1a-d). Two molecules with delta function-like PDFs were combined and the resolving power of maximum entropy deconvolution was investigated as a function of the mean distance between the states. With the molecule at 47 Ångströms as a reference, separations of -10 (Figure S1a), -4.6 (S1b), -2.3 (S1c) and +10 (S1d) Ångströms are presented, here blue lines represent the simulated mixtures, dashed red lines represent the expected result, and histograms in the 2nd row are results from CTB analysis. All PDFs presented in Figure S1 are created with an accuracy of 7%, thus the method should be able to resolve populations separated by more than ~3.3 Ångströms. Deconvoluted distributions resolve two states for each of the simulated binary distributions studied except the one with a separation of 2.3 Ångströms (Figure S1c), consistent with the theoretical resolution for maximum entropy deconvolution. Results of binary mixtures from CTB analysis are broad and relatively featureless, in stark contrast to the deconvoluted PDFs. Visually, only the -10 Å (Figure S1a) and +10 Å (S1d) histograms appear bimodal by CTB analysis and further interpretation of these distribution is possible only through Gaussian fitting. This demonstrates that the MIM analysis in conjunction with maximum entropy deconvolution has reached theoretical resolution limit and that the results are consistent in terms of weight, position and shape of the distribution.

Since typical data analysis is not done on single trajectories, intrinsic resolution of the two data analysis methods has also been investigated for simulated data sets comprised of a mixture of 200

experimental single-molecule trajectories re-sampled from poly-L-proline peptides of different lengths (Figure S1e-h). Here simulated distributions are created by combining poly-L-proline samples P_{15} and P_{24} in a 1:1 ratio (Figure S1e) as well as P_8 , P_{15} and P_{24} in ratios of 2:1:1 (S1f), 1:1:1 (S1g) and 1:1:2 (S1h). All ratios are calculated according to total trajectory length in each simulated data set. Although interpretation of the results of this exercise is less straight forward than the mixtures of two delta functions presented above, the deconvoluted PDFs are unambiguously able to resolve each population in the distribution, as well as reproduce approximately the correct weights and overall shape. By contrast the distributions produced by CTB are broad and although some of the information about the position and relative weight of states can be recovered by Gaussian fitting, the results obtained in this manner are non-optimal. Whereas the PDFs produced by MIM analysis have a meaningful shape, the CTB analysis contains little information about the overall shape of the distribution due to the broadening introduced by experimental noise.

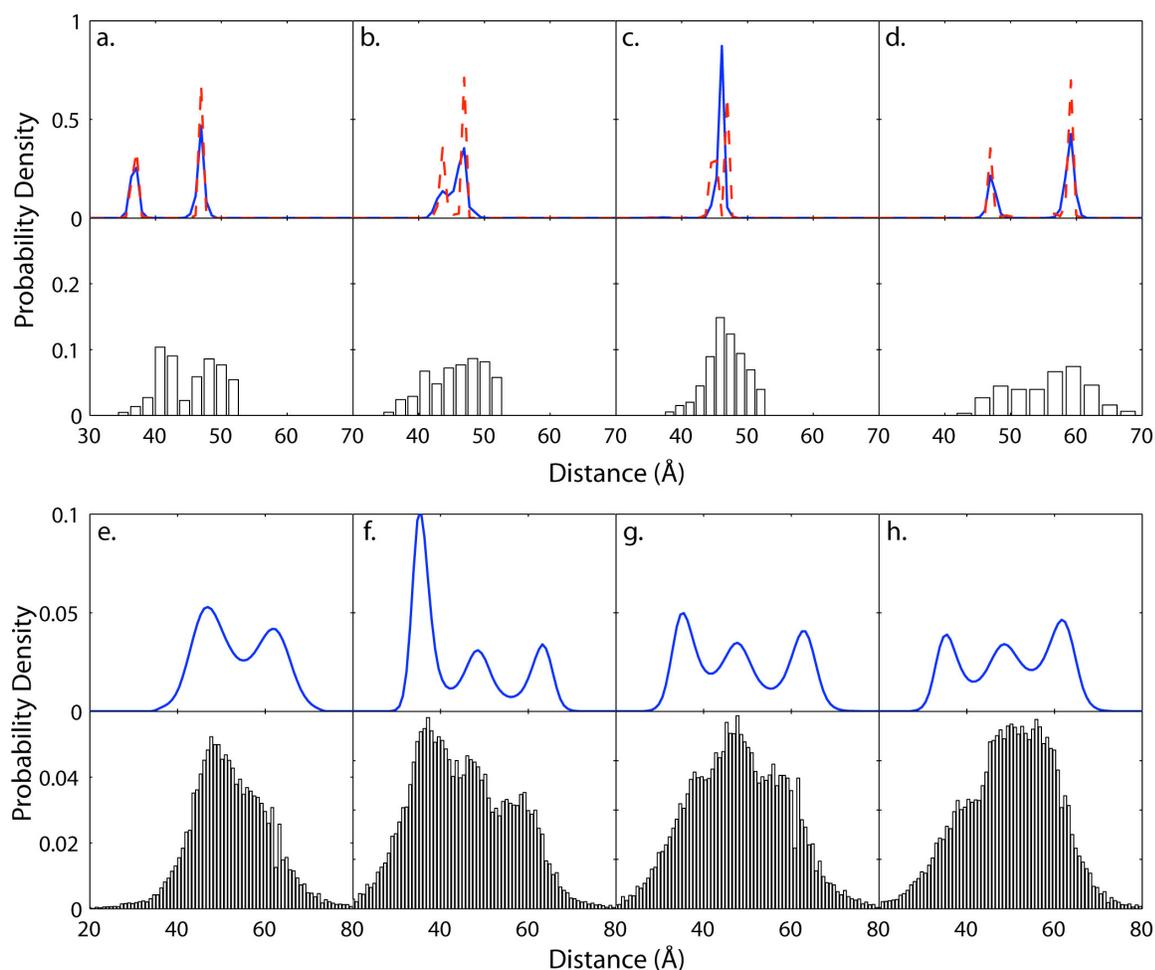


Figure S1 – Investigation of intrinsic resolution of single-molecule data analysis methods. Binary mixtures of two single molecule trajectories separated by -10 (a), -4.6 (b), -2.3 (c) and +10 (d) Ångströms, with the trajectory at 47 Ångströms as a reference.

Blue lines are the deconvoluted binary mixture from maximum information method (MIM) analysis, dashed red lines represent the expected result from individual trajectories and histograms in the 2nd column are the results from constant time binning (CTB) analysis. Deconvolution is expected to have a resolution of ~ 3.3 Ångströms, consistent with the results presented, whereas CTB analysis has an apparent resolution of 10 Ångströms. Simulated data sets consisting of mixtures of 200 poly-proline trajectories at ratios of 1:1 poly-proline- P_{15} : P_{24} (e) as well as mixtures of P_8 : P_{15} : P_{24} at 2:1:1 (f), 1:1:1 (g) and 1:1:2 (h). MIM results are in the top column while CTB results are in the bottom column. The MIM unambiguously reveals populations at the correct position and weight and contains meaningful information on the shape of the distribution, whereas the CTB results rely on fitting to Gaussian distributions for their interpretation and are significantly broadened due to photon-counting noise.

III. Single-molecule spectra and lifetime distributions

To ensure that the distance distributions measured in these experiments reflect actual lengths of poly-L-proline peptides in solution it is important to know the fluorescence characteristics of the attached donor and acceptor fluorophores when the peptide is immobilized on the coverslip surface. Deviations from the bulk fluorophore characteristics can skew the average single molecule length determination, and molecule-to-molecule variations distort the length distribution derived from single molecule measurements. This was first clearly demonstrated by Talaga *et al.* who showed that immobilization of a peptide by absorption to a silanized quartz surface caused artifacts in both conformation and dynamics when compared to the same peptide in solution [7]. In the present work, immobilization is achieved through a biotin-PEG passivated quartz cover slip to which the streptavidin-bound poly-L-prolines can be specifically attached. Though the use of PEG is thought to prevent non-specific interactions with the quartz substrate, careful control experiments must be performed to ensure that this is the case.

In order to show that the dye photo-physics are unaffected by proximity to the quartz or immobilization, dye emission spectra and excited-state lifetime were measured at the single-molecule level using a multi-parameter fluorescence microscope [8, 9]. Experiments were performed on doubly labeled poly-L-proline, P_{15} . Figure S2 shows a sample single-molecule intensity-versus-time trajectory from which spectra and lifetime data for both donor and acceptor have been extracted. Donor data is extracted from the region of the trajectory after the acceptor has bleached while acceptor data is measured from the region of the trajectory undergoing FRET. The measured single molecule donor and acceptor spectra can be well fit by the bulk emission spectra of the individual dyes as seen in Figure S2.

There is very little variation in the spectra between individual molecules. Representative fluorescence decay curves extracted from the single molecule trajectory are shown in Fig. S2 along with single exponential fits to the decays including convolution with the instrument response function. The lifetimes of both dyes are substantially longer when conjugated to the peptides than the reported lifetimes of the free dyes (Invitrogen). The results of these experiments confirm that the effect of the immobilization on the fluorophore characteristics is small and the single molecule distributions should be described by the bulk properties.

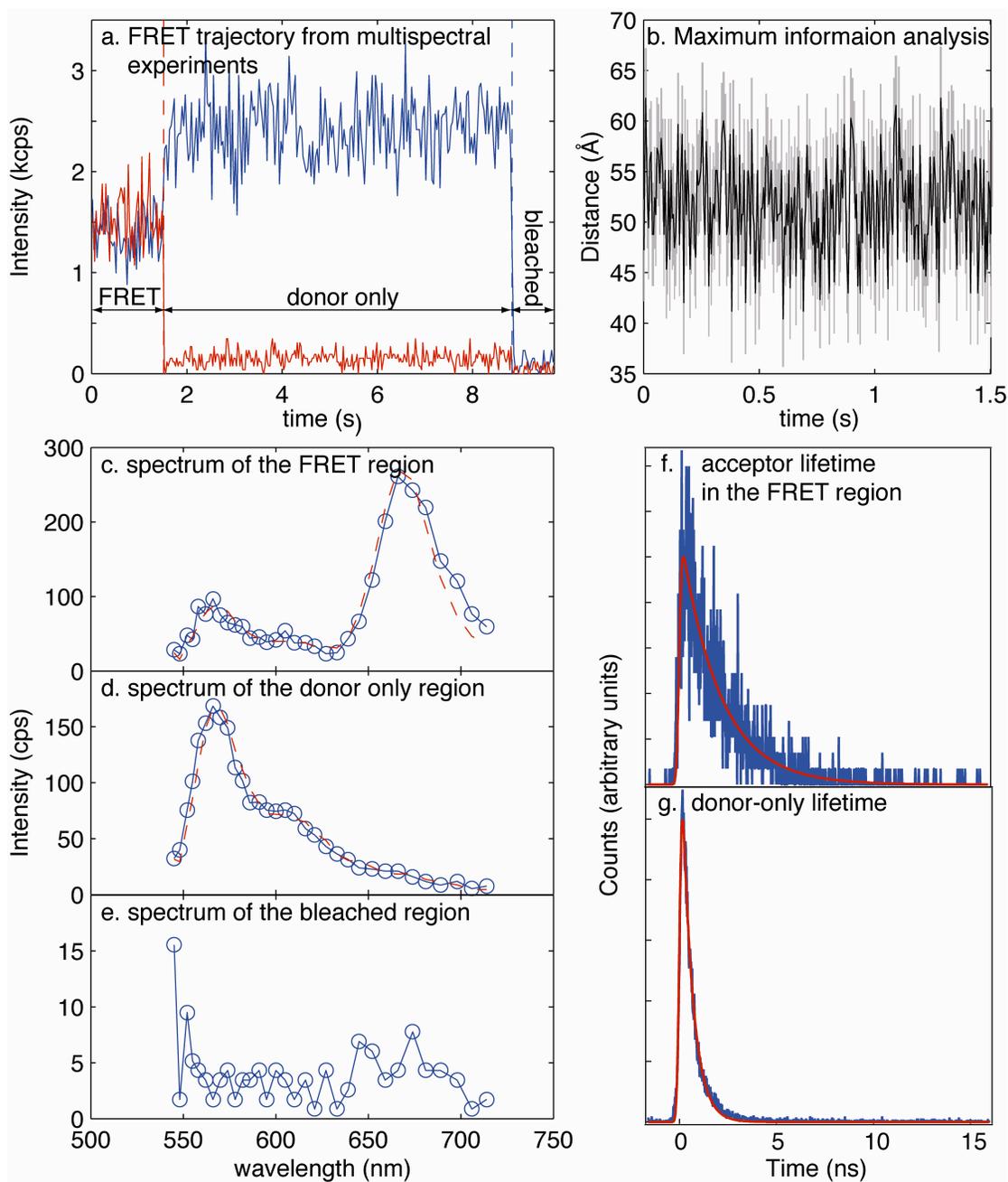


Figure S2 – Dye spectra and lifetime measured at the single-molecule level using a multispectral confocal microscope. A sample single-molecule trajectory is displayed in

panel (a). Acceptor emission is red while donor emission is blue. Dashed vertical lines represent the time at which each dye photo-bleaches. The maximum-information analysis of the FRET region is shown in panel (b). Spectral data for the FRET region of the trajectory is displayed in (c). Dashed red line represents a fit of the bulk spectra to the spectra obtained from single-molecule experiments. The donor spectrum obtained from the region of the trajectory after the acceptor dye has bleached is shown in (d) and also is in excellent agreement with the bulk counterpart. Acceptor lifetime information (f) is obtained over the wavelength range of 640-730 nm and is shown with the result of a 2 ns lifetime single exponential fit (in red) including convolution with the instrument response. The donor lifetime (g) is obtained from the region of the trajectory after the acceptor dye has bleached and is shown with the corresponding single exponential fit (lifetime 550 ps).

IV. Single-molecule polarization measurement of dye randomization

Accuracy of absolute distances measured in FRET experiments relies on orientational randomization of the dyes, manifested as κ^2 in the calculation of the Förster radius (R_0). This calculation assumes $\kappa^2 = 2/3$ for complete orientational randomization of the dyes thus deviation from this ideal case will result in misestimation of R_0 . It has been previously demonstrated that κ^2 approaches the $2/3$ theoretical limit assuming a randomization of dyes on a similar timescale to the inter-photon distance in single-molecule experiments [10, 11]. Single-molecule polarization modulation experiments have been used to demonstrate this assumption experimentally. Though, these results have been presented previously [1, 2], they are briefly summarized here. The plane polarized excitation source was modulated by 90° on the millisecond time scale with a Pockel's cell. Dyes with a freely randomized dipole should show constant emission intensity whereas dyes with a static or restricted dipole orientation should show polarization dependent emission intensity since the absorption of polarized light is dependent on the orientation of the dye dipole. An emission intensity dependence of polarized light direction would be manifested as intensity fluctuations on the same timescale as the Pockel's cell oscillation and can be readily detected as significant positive correlation in intensity auto-correlation function analysis [2]. Representative intensity auto-correlation functions for these experiments are presented for acceptor-only labeled poly-L-proline- P_{24} specifically immobilized via streptavidin-PEG chemistry (Figure S3b) as well as nonspecifically spin-coated on a glass cover slip (Figure S3a). Spin-coated samples have a static dipole orientation, which is manifested by significant correlation in their

intensity auto-correlation functions at the frequency of Pockel's cell oscillation (dashed red lines in Figure S3). By contrast, specifically immobilized samples show no correlation in their intensity auto-correlation functions indicating complete randomization of the dye on the experimental timescale. Polarization for the acceptor dye is presented since this dye is in closer proximity to the streptavidin and quartz substrate than the donor; however, similar results were obtained for polarization modulation experiments of the donor dye. Taken with the results of single-molecule spectral and lifetime measurements, these results demonstrate that dye properties are unaffected by the immobilization strategy used and that the Förster radius is consistent between bulk and single-molecule experiments.

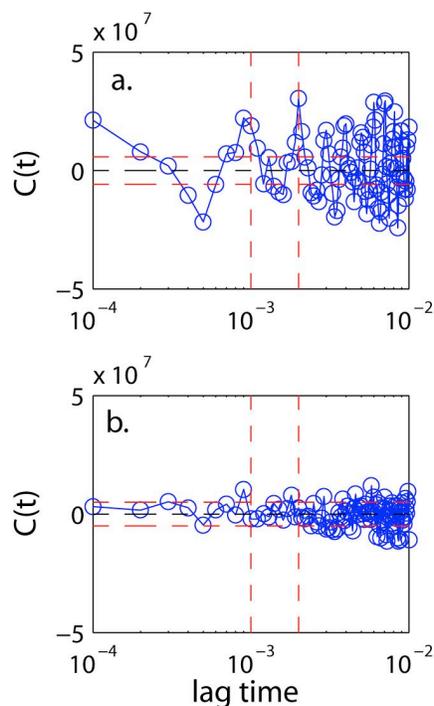


Figure S3 – Intensity auto-correlation functions of single-molecule polarization modulation experiments for acceptor-only labeled poly-L-proline- P_{24} . The polarization of the excitation source is modulated by 90° at 1 kHz with a Pockel's cell in order to demonstrate randomization of the dye orientation on the experimental timescale. (a) Spin coated proline on a glass surface showing significant auto-correlation due to a static dye orientation and (b) streptavidin immobilized proline on biotin-PEG passivated quartz showing no significant auto-correlation due to a randomized dye orientation. See text for further discussion. Vertical red lines are plotted at one and two times the Pockel's cell oscillation frequency to aid visualization. Error bars were calculated according to [2].

V. Steady-State FRET in trifluoroethanol

The solvent trifluoroethanol (TFE) is known to reduce the frequency of cis-isomers in poly-proline peptides [12, 13]. ensemble FRET experiments were conducted in TFE in order to demonstrate that the ensemble and single-molecule FRET measurements of poly-L-proline in aqueous buffer are consistent with the model in which cis-isomers exist in an otherwise trans helix. Figure S4 shows the ensemble FRET data in TFE (green line) compared to the ensemble FRET data in aqueous buffer (red) and the mean distances from single-molecule experiments (blue). As in the main text Figure 6, the dashed black line represents the end-to-end distance expected from a perfectly rigid all-trans proline using a rise of 3.12 Å per proline residue derived from the crystal structure [14]. The distances in TFE are consistently longer than those found in aqueous buffer as predicted. The error bars for poly-proline-P₈ are large since this sample had a FRET efficiency of 0.99, thus the distance measurement is expected to be less accurate [5]. An R₀ value of 65 Å was used for the FRET dye pair in TFE compared to a value of 51 Å in aqueous buffer. This value was calculated assuming the only changes in R₀ were due to refractive index and quantum yield according to the relationship $R_{0,TFE} = \left(R_{0,ref}^6 \Phi_{D,TFE} / \Phi_{D,ref} n_{ref}^4 / n_{TFE}^4 \right)^{1/6}$ where $\Phi_{D,TFE}$ is the fluorescence quantum yield of the donor in TFE and n is the refractive index. The quantum yield ratio is calculated by measuring the relative molar fluorescence in the two solvents. The refractive index for water is $n_{ref}=1.3333$ while that for TFE is $n_{TFE}=1.2907$ [15].

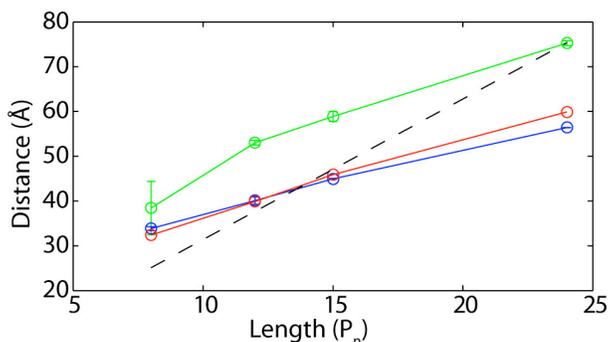


Figure S4 – Comparison of steady-state bulk FRET in TFE (green line) and aqueous buffer (red line) to mean distances from single-molecule experiments (blue curve). For comparison, the predicted end-to-end distances for perfectly rigid poly-L-prolines have been included (dashed black line); no correction has been made for additional distance contributions expected from dyes and dye-linkers.

VI. The freely jointed chain model

The freely jointed chain (FJC) model has been widely used to model polymers, denatured proteins, unstructured peptides and peptidyl segments, and so forth. In this model, a polymer chain is considered as composed of M rigid segments of length l joined together in which each segment can freely move around about the joint. The probability density function for finding a polymer of contour length $l_c \equiv Ml$ exhibiting an end-to-end distance of R is [16],

$$p(R) = 4\pi R^2 \left(\frac{2\pi M l^2}{3} \right)^{-3/2} \exp\left(-\frac{3R^2}{2Ml^2} \right). \quad (\text{S1})$$

The mean distance is therefore,

$$\langle R \rangle^{(\text{FJC})} = \int_0^\infty R p(R) dR = \sqrt{\frac{8Ml^2}{3\pi}} = \sqrt{\frac{8l_c l}{3\pi}}. \quad (\text{S2})$$

To fit the FJC model to the experimentally measured mean end-to-end distance, we consider the averaged effective length of the dye linkers ($L_{\text{linker}}^{(\text{FJC})}$) added to the overall distance. The contour length of a poly-L-proline of n residues was estimated by $l_c = n \times 3.12 \text{ \AA}$, where the increment per residue for the all-*trans* poly-L-proline II structure, 3.12 \AA , was deduced from the crystal structure by Cowan and McGavin [14]. We used the `fminsearch` function in Matlab (R2007a) to fit $L_{\text{linker}}^{(\text{FJC})}$ and l in Eq. (S2) for the following equality:

$$R^{(\text{expt})} = \langle R \rangle^{(\text{FJC})} + L_{\text{linker}}^{(\text{FJC})}. \quad (\text{S3})$$

Using data from Ref. [10], we found $\hat{L}_{\text{linker}}^{(\text{FJC})} = 1.47 \text{ \AA}$ and $\hat{l} = 47.22 \text{ \AA}$. The comparison of fitted result with data is displayed as the green line in the inset of Fig. 4 in the main text. The optimal fitting parameters ($\hat{L}_{\text{linker}}^{(\text{FJC})}$ and \hat{l}) were also used in Eq. (S1) to produce the distance distribution functions shown in Fig. 5 in the main text.

VII. The self-avoiding chain model.

The self-avoiding chain (SAC) model is a more physically reasonable refinement of the FJC model, which takes into account excluded volume effects from the other monomers of the chain. Though this model has no closed form solution the mean end to end distance, $\langle R \rangle^{\text{SAC}}$, is known to obey the scaling relationship [17],

$$\langle R \rangle^{SAC} = \left(AN^{2\nu} \left(1 + \frac{B}{N^\Delta} + \frac{C}{N} + \dots \right) \right)^{1/2} \quad (S4)$$

where N is the number of monomers, ν and Δ are critical exponents with values of 0.588 and 0.56 respectively and A , B , and C are model dependent constants. The distribution of end-to-end distances is show by des Cloizeaux to be [18],

$$p(r) = \frac{a}{\langle R \rangle^{SAC}} \left(\frac{r}{\langle R \rangle^{SAC}} \right)^{2+\theta} \exp \left(-b \left(\frac{r}{\langle R \rangle^{SAC}} \right)^\delta \right) \quad (S5)$$

where $\langle R \rangle^{SAC}$ is the end-to-end distance from Eq. (S4), θ and δ have values of 0.2753 and 2.427 respectively [17] and a and b are normalization constants whose values satisfy the relationship $\int p(r)dr = \int r^2 p(r)dr = 1$.

Eq. (S4) was used to fit the mean experimental data in the same manner as the FJC model by comparing the mean fret distances to,

$$R^{\text{expt}} = \langle R \rangle^{SAC} + L_{\text{linker}}^{SAC}. \quad (S6)$$

The best-fit parameters were found to be $A = 338 \text{ \AA}^2$, $B = 1.63 \times 10^{-8}$, $C = 3.63 \times 10^7$, and $L_{\text{linker}}^{SAC} = 15.08 \text{ \AA}$. The normalization parameters in Eq. (S5) were found to be $a = 3.766$ and $b = 1.272$.

VIII. The worm-like chain model

The worm-like chain (WLC) model treats polymers as semi-flexible rods and has been used by Schimmel and Flory to describe the nature of poly-L-proline [19]. An analytical form for the end-to-end distance probability density of a WLC polymer has been found by Becker *et al.* [20], which reads,

$$p(r) = J_{\text{SYD}} \left(\frac{1-cr^2}{1-r^2} \right)^{5/2} \exp \left(\frac{\sum_{i=1}^0 \sum_{j=1}^3 c_{i,j} \kappa^i r^{2j}}{1-r^2} \right) \exp \left(\frac{d\kappa ab(1+b)r^2}{1-b^2r^2} \right) I_0 \left(-\frac{d\kappa a(1+b)r}{1-b^2r^2} \right), \quad (S7)$$

where $a = 14.054$, $b = 0.473$, and

$$J_{\text{SYD}} = \begin{cases} 112.04\kappa^2 e^{0.246/\kappa - a\kappa}, & \kappa > 1/8, \\ (3/4\pi\kappa)^{3/2} e^{-3r^2/4\kappa}, & \kappa \leq 1/8, \end{cases}$$

$$c = 1 - \left(1 + (0.38\kappa^{-0.95})^{-5} \right)^{-1/5},$$

$$d = \begin{cases} 1, & \kappa < 1/8, \\ 1 - \left(0.177 / (\kappa - 0.111) + 6.40 (\kappa - 0.111)^{0.783} \right)^{-1}, & \kappa \geq 1/8, \end{cases}$$

$r = R/l_c$, $\kappa = l_c/l_p$ with $l_c = n \times 3.12 \text{ \AA}$ being the contour length of the polymer and l_p the persistence length. The expectation value of end-to-end distance was calculated by numerical integration of the radial distribution function, $\langle R \rangle^{(WLC)} = \int R p(r) l_c dR = \int R p(R) dR$.

Similar to the FJC model, Eq. (S7) was used to fit Eq. (S8) to the experimentally measured mean end-to-end distances shown in Fig. 4 in the main text,

$$R^{(\text{expt})} = \langle R \rangle^{(WLC)} + L_{\text{linker}}^{(WLC)}, \quad (\text{S8})$$

where the adjustable parameters are the persistence length l_p and the linker length $L_{\text{linker}}^{(WLC)}$. The optimal fitting parameters are $\hat{L}_{\text{linker}}^{(WLC)} = 8.44 \text{ \AA}$ and $\hat{l}_p = 13.74 \text{ \AA}$. These optimized parameters were then used in Eq. (S8) to produce the red WLC distributions shown in Fig. 5 in the main text.

The numerical value of the fitted \hat{l}_p is similar to the persistence length of 23 \AA obtained by Mattice and Mandelkern from bulk viscoelastic measurements [21, 22]. Indeed, in the previous work [10], $l_p = 23 \text{ \AA}$ was used to constrain the fitting, for which very good results were also obtained. Sahoo *et al.* have made a comprehensive comparison of short ($n = 1 - 6$) poly-L-proline results from different research groups and essentially obtained excellent fits for both the FJC (Gaussian model in their article) and the WLC models [23]. They have concluded that the persistent length is about $30 - 70 \text{ \AA}$. Doose *et al.* have also studied the end-to-end distances of short poly-L-prolines ($n = 1 - 10$) using excited-state electron transfer. They have provided strong evidence that *cis* residues exist in the otherwise *trans* polymers. An independent work by Best *et al.* have also provided strong NMR evidence that *cis* configurations do exist in poly-L-prolines in water, albeit to a small yet to be determined extent [13]. In the concluding paragraph, Soose *et al.* have cautioned the applicability of the WLC model, which applies to semi-flexible rod but not segmented repeats. These results underscore the conflicting views regarding the nature of poly-L-proline, which has served as an important model for poly-peptide configurations. The inability to discriminate among different models further emphasizes the difficulty in resolving models (including those based on the same basic physical picture but having different persistent lengths) based on the apparent mean distances along.

In Fig. 6 in the main text, Eq. (S7) was used to fit the experimentally obtained end-to-end distribution functions in which the effective persistence length l_p and the apparent contour length L were the adjustable parameters.

IX. References

- [1] J.A. Hanson, K. Duderstadt, L.P. Watkins, S. Bhattacharyya, J. Brokaw, J.W. Chu, H. Yang, Proc. Natl. Acad. Sci. U.S.A., 104 (2007) 18055-18060.
- [2] J.A. Hanson, H. Yang, J. Chem. Phys., 128 (2008) 214101.
- [3] J.A. Hanson, H. Yang, J. Phys. Chem. B, 112 (2008) 13962-13970.
- [4] L.P. Watkins, H. Yang, J. Phys. Chem. B, 109 (2005) 617-628.
- [5] L.P. Watkins, H. Yang, Biophys. J., 86 (2004) 4015-4029.
- [6] R. Roy, S. Hohng, T. Ha, Nat. Methods, 5 (2008) 507-516.
- [7] D.S. Talaga, W.L. Lau, H. Roder, J.Y. Tang, Y.W. Jia, W.F. DeGrado, R.M. Hochstrasser, Proc. Natl. Acad. Sci. U.S.A., 97 (2000) 13021-13026.
- [8] A.K. Luong, G.C. Gradinaru, D.W. Chandler, C.C. Hayden, J. Phys. Chem. B, 109 (2005) 15691-15698.
- [9] C.S. Xu, H. Kim, H. Yang, C.C. Hayden, J. Am. Chem. Soc., 129 (2007) 11008-11009.
- [10] L.P. Watkins, H.Y. Chang, H. Yang, J. Phys. Chem. A, 110 (2006) 5191-5203.
- [11] H. Yang, Israel J. Chem., 49 (2009) 313-322.
- [12] D.S. Clark, J.J. Dechter, L. Mandelkern, Macromolecules, 12 (1979) 626-633.
- [13] R.B. Best, K.A. Merchant, I.V. Gopich, B. Schuler, A. Bax, W.A. Eaton, Proc. Natl. Acad. Sci. U.S.A., 104 (2007) 19064-19066.
- [14] P.M. Cowan, S. McGavin, Nature, 176 (1955) 501-503.
- [15] CRC Handbook of Chemistry and Physics 2008-2009, in: D.R. Lide (Ed.), CRC Press, New York, 2008, pp. Section 3.
- [16] P.C. Hiemenz, Polymer Chemistry: The Basic Concepts, Marcel Dekker, Inc., New York, 1984.
- [17] J.P. Valleau, J. Chem. Phys., 104 (1996) 3071-3074.
- [18] J. des Cloizeaux, Phys. Rev. A, Gen. Phys., 10 (1974) 1665-1669.
- [19] P.R. Schimmel, P.J. Flory, Proc. Natl. Acad. Sci. U.S.A., 58 (1967) 52-59.
- [20] N.B. Becker, A. Rosa, R. Everaers, Eur. Phys. J. E, 32 (2010) 53-69.
- [21] W.L. Mattice, L. Mandelkern, J. Am. Chem. Soc., 93 (1971) 1769-1777.
- [22] W.L. Mattice, L. Mandelkern, Biochemistry, 9 (1970) 1049-1058.
- [23] H. Sahoo, D. Roccatano, A. Hennig, W.M. Nau, J. Am. Chem. Soc., 129 (2007) 9762-9772.